Human htFIIIA gene and coded htfIIIA protein

The present invention relates to the gene coding for the human transcription factor hereafter called htFIIIA (or htfC2) gene and the coded htfIIIA protein, as well as the use of this htFIIIA gene and that of the coded htfIIIA protein in the diagnosis and identification of certain diseases related to the transcription mechanism.

Hereafter the gene coding for the transcription factor TFIIIA will be called tfIIIA (or tfC2) and the gene coding for the human transcription factor htfIIIA will be called htfIIIA.

The human htFIIIA gene codes therefore for the corresponding htFIIIA protein.

We will also use the following abbreviations below: AA for amino acids, NA for nucleic acids, bp for base pairs, DNA for deoxyribonucleic acid, cDNA for complementary DNA, RNA for ribonucleic acid, RNase for ribonuclease and C for deoxycytidine.

The term screening which indicates a specific screening technique and the term primer which indicates an oligonecleotide used as primer will also be used.

The tfIIIA gene and the corresponding tfIIIA protein are involved in the regulation of the biological transcription mechanism as indicated below.

Since the tfIIIA protein was purified as transcription factor for the first time in 1980 from Xenopus ovocytes [Segall et al, J. Biol. Chem., 255, 11986-11991 (1980)], work has been carried out in vivo and in vitro within the Xenopus to study the mechanism of transcription control exercised by TFIIIA. It has thus been shown that Xenopus TFIIIA is necessary for the initiation of the transcription of 5S RNA gene [Sakonji et al, Cell 19, 13-25 (1980)] and binds to a internal control region of the 5S RNA gene [Bogenhagen et al, Cell, 19,27-35 (1980)].

The nucleotide sequence of the cDNA of Xenopus TFIIIA and the corresponding amino acid sequence have already been published [Ginberg et al, Cell 39,479-489 (1984)]. It can be

noted that this gene codes for a structure of 9 zinc fingers, a zinc finger corresponding to the repetition of the CYS2 HIS2 (C2H2) moiety. This zinc finger structure is considered an essential domain for a group of proteins which bind themselves to the DNA (DNA binding proteins) [Miller et al, Embo J., 4, 1607-1614 (1985)].

In this way tránscription factors in human beings, binding to the DNA which also have this zinc finger structure such as for example XT1 of the Wilms human tumor gene, [Gessier et al, Nature, 343, 774-778 (1990)], the YY1 human transcription repressor [SHI et al, Cell, 67, 377-388 (1991)], the MAZ protein combined with the human MYC gene [Bossone et al, Proc. Natl. Acad. Sci., USA, 89, 7452-7456 (1992)] or also spl [Kuwahara et al, J.Biol. Chem., 29, 8627-8631 (1990)] are known.

Studies have been carried out in order to isolate the human htFIIIA gene, but until now none have led to discovery of the true sequence of the htFIIIA gene.

On one hand the studies described in the European Application EP 0704526 (Fujisawa et al), can thus be mentioned and are examined in the article: Arakawa et al (1995), Cytogenet Cell Genet 70, 235-238, which have led to a sequence that we will call Arakawa htfIIIA and on the other hand the studies described in the article: DREW et al (1995), Gene 159, 215-218, which have led to a sequence that we will call DREW htfIIIA. These DREW and ARAKAWA htfIIIA sequences are represented in Figures 4 and 5 respectively below. The documents indicated above therefore each describe a sequence of the htfIIIA gene but these two sequences differ from one another by a few nucleotides and differ from the htfIIIA gene of the present Application as indicated below.

The present invention has made it possible to isolate the gene coding for the human transcription factor hTFIIIA.

The present invention has also made it possible to reveal the nucleic acid sequence of the htfIIIA gene and also the amino acid sequence of the hTFIIIA protein coded by this gene.

Therefore a subject of the present invention is the DNA

sequence of the htfIIIA gene coding for a protein having the biological function of human transcription factor htfIIIA.

A precise subject of the present invention is the DNA sequence of the htfIIIA gene of human transcription factor hTFIIIA as defined above, coding for the amino acid sequence SEQ ID N°2.

Such a SEQ ID n°2 sequence of the present invention therefore comprises 365 amino acids.

A subject of the present invention is also the DNA sequence of the htfIIIA gene as defined above, containing the nucleotide sequence SEQ ID N°3.

A subject of the present invention is the DNA sequence of the htfIIIA gene as defined above, containing the nucleotide sequence SEQ ID N°4.

A subject of the present invention is also the DNA sequence of the htfIIIA gene as defined above, corresponding to the nucleotide sequence SEQ ID N°3.

The sequence SEQ ID N°3 therefore comprises 1273 nucleotides.

A particular subject of the present invention is the DNA sequence of the htfIIIA gene as defined above, corresponding to the nucleotide sequence SEQ ID N°4. The sequence SEQ ID N°4 therefore comprises 1213 nucleotides.

The sequence SEQ ID N°1 represents the nucleotide sequence of the htFIIIA gene on the upper line according to the present invention i.e. SEQ ID N°3, and the corresponding amino acid sequence (AA) of this nucleotide sequence i.e. SEQ ID N°2 on the lower line.

Figures 1 and 2 below represent the AA sequence coded by htfIIIA of the present invention SEQ ID N°2 on the upper line, and the AA sequences coded by the DREW htfIIIA genes, in Figure 1, and ARAKAWA genes in Figure 2 on the lower line respectively, these DREW and ARAKAWA sequences are as published in the documents referred to above.

Figure 3 below represents the comparison of AA sequences coded by the DREW and ARAKAWA htfIIIA genes respectively with the AA sequence coded by Arakawa htfIIIA on the upper line and the AA sequence coded by DREW htfIIIA on the lower line.

Figure 2 therefore shows, that the corresponding AA

sequence of htfIIIA according to the present invention comprises differences from the AA sequence published in the ARAKAWA article or EP 0704 526, in particular in the corresponding positions 105 and 163, 156 and 214, 320 to 329 and 378 to 387 respectively, these positions being given in relation to the numbering indicated in Figure 2.

Figure 2 also shows that the AA sequence coded by htfIIIA of the present invention begins at position 59 of the AA sequence of Arakawa htfIIIA.

Figure 3 shows that the AA sequences coded by Arakawa and DREW htfIIIA comprise differences at the corresponding positions 214 and 154, 378-387 and 318-327 respectively, these positions being given in relation to the numbering indicated in Figure 3.

Figure 5 shows that the Arakawa htfIIIA sequence codes for a protein, the amino acid sequence of which, indicated in EP 0704 526, begins with the AA methionine specified by the ATG codon which is found in position 20-22 and the translation stops at a TAA codon. If the nucleotide sequence of htfIIIA according to the present invention SEQ ID N°3 is compared with the nucleotide sequence of EP 0704 526 i.e. Arakawa htfIIIA shown in Figure 5 (sequence p11-12-13 of EP 0704 526), it can be noted that it lacks a C nucleotide in position 127 of the EP 0704 526 sequence. This additional C nucleotide results in a shift in the translation of amino acids of this nucleotide sequence: in fact, the ATG which is found in position 20-22 of the ARAKAWA sequence shown in Figure 5 and which is considered to be a start codon of proteinic synthesis by ARAKAWA, is therefore no longer in the same reading frame because of this shift. By taking into consideration this additional C nucleotide, the translation of AA reveals a TGA stop codon in position 57-59 of the ARAKAWA sequence shown in Figure 5. Consequently, the start codon of proteinic synthesis according to the present invention is located downstream of this stop codon. Translation experiments in vitro of SEQ ID N°4 and expression tests in mammalian cells such as Cos cells have made it possible to identify the start codon of the proteinic

synthesis of hTFIIIA according to the present invention.

This start codon of proteinic synthesis of hTFIIIA according to the present invention is the CTG codon in position 176-178 of SEQ ID N°3 (which would correspond to position 194-196 of the ARAKAWA sequence shown in Figure 5).

The coding section of the htFIIIA gene of the present invention begins therefore with this CTG codon which is found in position 176-178 of SEQ ID N°3 which should correspond to the AA Leucine and which in fact corresponds to the AA Methionine as this codon is recognised as a start codon (ref: David S. Peabody The Journal of Biological Chemistry, vol. 264, n°9, pp. 5031-5035, 1989).

Consequently, as Figure 2 shows, the ARAKAWA hTFIIIA protein is longer than the hTFIIIA protein of the present invention.

Furthermore, if the hTFIIIA protein of the present invention and the DREW hTFIIIA protein are compared (comparison shown in Figure 1), it is noticed that the amino acid threonine in position 105 of the hTFIIIA protein of the present invention corresponds to an asparagine residue in position 103 in the DREW hTfIIIA sequence and that the two first AA, M and D of the hTFIIIA protein of the present invention have not been determined for the DREW hTFIIIA protein. The absence of codons specifying these AA and in particular the absence of the start codon of proteinic synthesis, does not permit the expression of this protein. The DREW htfIIIA sequence shown in Figure 4 is therefore incomplete, and this is recognised by the authors of the publication referred to above (DREW et al on page 216 lines 39-41). It can be noted moreover, that the authors of this article also think that the start codon of the DREW htfIIIA sequence should correspond to a methionine coded by ATG as in the ARAKAWA sequence.

The htfIIIA gene according to the present invention is therefore different from the DREW and ARAKAWA htfIIIA genes (EP 0704526) and codes for a hTFIIIA protein, the AA sequence of which is different from that of the DREW and ARAKAWA hTFIIIA proteins.

Therefore a particular subject of the present invention is the DNA sequence of the htfIIIA gene as defined above containing the nucleotide sequence SEQ ID N°3.

A more particular subject of the present invention is the DNA sequence as defined above having the sequence beginning at nucleotide 176 and finishing at nucleotide 1270 of SEQ ID N°3.

One such sequence of the present invention therefore begins at a CTG codon and thus comprises 1095 nucleotides.

A subject of the present invention is the DNA sequence coding for the human transcription factor hTFIIIA as defined above as well as the DNA sequences which hybridize with it and/or show a significant homology with this sequence or fragments of it and coding for proteins having the same function.

By sequences which hybridize are included DNA sequences which hybridize with one of the DNA sequences above under standard conditions of high, medium or low stringency. By proteins with the same function are included polypeptides with the same transcription factor function. The stringency conditions are those carried out in conditions known to a person skilled in the art, such as those described by Sambrook et al (1989) Molecular cloning, Cold Spring Harbor Laboratory Press, 1989. Such stringency conditions are for example hybridization at 65°C, for 18 hours in a 5 x SSPE; 10 x Denhardt's; 100µg/ml ssDNA; 1 % SDS solution followed by washing 3 times for 5 minutes with 2 x SSC; 0.05 % SDS, then washing 3 times for 15 minutes at 65°C in 1 x SSC; 0.1 % SDS. The high stringency conditions for example include hybridization for 18 hours at 65°C in a 5 x SSPE; 10 x Denhardt's; 100µg/ml ssDNA; 1 % SDS solution, followed by washing twice for 20 minutes with a 2 x SSC; 0.05 % SDS solution at 65°C followed by a final wash for 45 minutes in a 0.1 x SSC; 0.1 % SDS solution at 65°C. Medium stringency conditions for example include a final washing for 20 minutes in a 0.2 x SSC, 0.1 % SDS solution at 65°C.

By sequences which show a significant homology are included sequences with a nucleotide sequence with a similarity of at

least 50 % with one of the DNA sequences above and which codes for a protein having the same transcription factor function.

A subject of the present invention is also the DNA sequence as defined above comprising modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein having the same biological activity as the human transcription factor htfIIIA.

A particular subject of the present invention is the DNA sequence as defined above as well as similar DNA sequences which have a nucleotide sequence homology of at least 50 % or at least 60 % and preferably at least 70 % with the said DNA sequence.

Therefore a subject of the present invention is also the DNA sequence as defined above as well as the DNA sequences which code for a protein, the AA sequence of which has a homology of at least 40 % and in particular of 45 % or at least 50 %, rather at least 60 % and preferably at least 70 % with the AA sequence coded by the said DNA sequence.

The gene of the present invention is represented as a single strand DNA sequence but it is understood that the present invention includes the complementary DNA sequence of this single strand DNA sequence, and also includes the so-called double strand DNA sequence constituted by these two DNA sequences complementary to each other.

The DNA sequence of the present invention is an example of the combination of codons coding for the amino acids corresponding to the amino acid sequence SEQ ID N°2, but it is also understood that the present invention includes any other arbitrary combination of codons coding for this same amino acid sequence SEQ ID N°2.

The DNA sequence as defined above or this modified DNA sequence as indicated above, can be prepared according to techniques known to a person skilled in the art and in particular those described in the book by Sambrook, J. Fritsh, E. F. § Maniatis, T. (1989) entitled: " Molecular cloning: a laboratory manual ", Laboratory, Cold Spring Harbor NY. In particular the DNA sequence above can be a

cDNA sequence obtained by identification of the 3' and 5' parts of the coding sequence, then amplification of these parts using a DNA polymerase such as pfu polymerase or other DNA polymerases. The introduction, into the oligonucleotide sequence used for PCR, of restriction sites such as Hind III or SmaI allow the cloning of these fragments in appropriate vectors and then the restoration of the sought complete sequence. A detailed description of the operating conditions in which the present invention was carried out is given below.

A quite particular subject of the invention is the polypeptide having the function of human transcription factor hTFIIIA and having the amino acid sequence SEQ ID N°2 coded by the DNA sequence as defined above and the analogues of this polypeptide.

By analogues is understood the polypeptides the amino acid sequence of which has been modified by substitution, suppression or addition of one or more amino acids but which retain the same biological function. Such polypeptide analogues can be produced spontaneously or can be produced by post-transcriptional modification or also by modification of the DNA sequence of the present invention as indicated above, by using techniques known to a person skilled in the art: amongst these techniques, the directed mutagenesis technique (Kramer, W., et al., Nucl. Acids Res., 12, 9441 (1984); Kramer, W. and Fritz, H.J., Methods in Enzymology, 154, 350 (1987); Zoller, M.J. and Smith, M. Methods in Enzymology, 100.468 (1983)) can in particular be mentioned.

Modified DNA synthesis can also be carried out by using well-known chemical synthesis techniques such as the phosphotriester method for example [Letsinger, R.L and Ogilvie, K.K., K. Am. CHEM. Soc., 91.3350 (1969); Merrifield, R.B., Sciences, 150, 178 (1968)] or the phosphoamidite method [Beaucage, S.L and Caruthers, M .H., Tetrahedron Lett., 22, 1859 (1981); McBRIDE, L.J. and Caruthers, M.H. Tetrahedron Lett., 24 245 (1983)] or also by the combination of these methods.

The polypeptides of the present invention can therefore

be prepared by techniques known to a person skilled in the art, in particular partially by chemical synthesis or also by cDNA synthesis by expression in a procaryotic or eucaryotic host cell as indicated below.

A particular subject of the present invention is the process for the preparation of the recombinant htFIIIA protein having the amino acid sequence SEQ ID N°2. This process includes the expression of the DNA sequence as defined above in an appropriate host, then isolation and purification of the said recombinant protein.

To produce the polypeptide of the present invention, recombinant DNA techniques using genetic engineering and cell culture methods known to a person skilled in the art can in particular be used. The following stages can therefore be carried out: firstly preparation of the appropriate gene, then incorporation of this gene into a vector, transfer of the gene carrier vector into an appropriate host cell, expression of the gene and finally purification of the protein coded by this gene.

The DNA sequences according to the present invention and in particular SEQ ID N°3 or SEQ ID N°4 can be prepared according to techniques known to a person skilled in the art, in particular by chemical synthesis, by screening of a gene bank or a cDNA bank using oligonucleotide synthesis probes using known hybridization techniques or also by reverse transcriptase from messenger RNA (mRNA).

The advantage of the technique comprising firstly the isolation of mRNA by extraction of the total RNA then the synthesis of cDNA from this mRNA by reverse transcriptase particularly rests on the fact that the mRNA does not contain introns even though these non-coding sequences are present in the genomic DNA.

The following procedure can in particular be carried out. Firstly the total RNA originating from a cell line such as for example the Raji cell line (RNA Plus, BIOPROBE) is extracted, and from this RNA, synthesis of the sought cDNA is then carried out, in particular by using a kit such as the RNA PCR kit (Perkin Elmer).

It can be noted that within the scope of the present invention, two oligonucleotides located at the ends of the htfIIIA coding sequence published by ARAKAWA (Figure 5) were synthesized i.e. OLT5 and OLT3 and are defined as follows:

- OLT5: 5' CGGGGTACCAAAA ATG CGC AGC AGC GGC GCC GAC 3' i.e. SEQ ID N°5 and
- OLT3: 5' CGGTCTAGA TTA GCC AAG GGT AAG TAC TGC 3' i.e. SEQ ID N°9

but these two oligonucleotides have not made it possible to obtain an amplification product by PCR.

Thus, within the scope of the present invention, the hTFIIIA coding sequence was isolated in two stages: firstly identification of the 3' part then identification of the 5' part.

After identification of the 3' and 5' parts, a HindIII restriction site located on each of these fragments then made it possible to restore the complete sought sequence as indicated below in the experimental part.

The following process was then carried out:

The 3' part was amplified using pfu polymerase (STRATAGENE) using the OLT5.2 and TFIIIA 3'SmaI oligonucleotides as primer i.e.:

- OLT5.2: 5'TCCTTCCCTGACTGCAGCGCC 3' or SEQ ID N°6 and
- TFIIIA3'SmaI: 5'CCT CCC GGG GCC AAG GGT AAG TAC TGC AAC 3' or SEQ ID N°10

The amplification primers are chosen as a function of the part to be amplified according to the usual criteria of a person skilled in the art.

The primers used in the present invention were chosen in the Arakawa htfIIIA sequence shown in Figure 5.

The sequences SEQ ID N°6, SEQ ID N°7 and SEQ ID N°8 are located in positions 320-340 (5'→3'), 361-380 (reverse and complementary sequence) and 391-410 (reverse and complementary sequence) respectively of this Arakawa htfIIIA sequence.

The sequences SEQ ID N°5, SEQ ID N°9 and SEQ ID N°10 are located in positions 20-40 (5'→3'), 1271-1291 (reverse and complementary sequence) and 1268-1288 (reverse and

complementary sequence) respectively of this Arakawa htfIIIA sequence.

It can be noted that sequences SEQ ID N°5, SEQ ID N°9 and SEQ ID N°10 contain sequences corresponding to the restriction enzyme sites i.e. KpnI, XbaI and SmaI respectively.

The oligonucleotide TFIIIA 3' SmaI introduces a restriction site SmaI downstream of the coding sequence. This site permits, if necessary and if required, the fusion of the coding sequence for hTFIIIA with a coding sequence for a hemaglutinin epitope peptide designated " TAG HA ". The expression of the coding sequence for TFIIIA can therefore be combined with that of the coding sequence for TAG HA which can be detected by Western blot analysis, if the fusion gene is expressed.

For identification of the 5' part, this region was isolated by the 5' anchored PCR (5 race System, GIBCO BRL; pfu polymerase, STRATAGENE) technique. Two successive PCR's were carried out using the following oligonucleotides as primer: UAP and TFIIIAPCR5' for the first PCR and UAP and TFIIIA SEQ2 for the second PCR.

UAP is an oligonucleotide provided in the kit.

These oligonucleotides have the following sequences:

- TFIIIAPCR5': 5' CACAAACAAATGGTCTCC 3' or SEQ ID N°8
- TFIIIA SEQ2: 5' TGCACAGGTGCGCGTCAAGC 3' or SEQ ID N°7.

The products of these PCR's i.e. the amplified 5' and 3' fragments are then purified on agarose gel and cloned using the TA cloning kit (INVITROGEN). Sequencing is then carried out: the plasmid DNA of several independent clones is prepared (QIAGEN Plasmids KIT) and the fragments corresponding to the coding sequence of hTFIIIA are sequenced on the two strands (ABI 377XL sequencer, PERKIN ELMER).

The following process can then be carried out according to usual cloning techniques known to a person skilled in the art and in particular cloning by insertion of each fragment into a plasmid provided with the commercial kit (TA cloning Kit Invitrogen), then transformation of a bacterial strain by the plasmid thus obtained is then carried out. The XL1 Blue E. coli strain can in particular be used.

The clones are then cultured in order to extract the plasmid DNA according to standard techniques known to a person skilled in the art referred to above (Sambrook, Fritsh and Maniatis).

Sequencing of the DNA of the amplified fragment contained in the plasmid DNA is carried out.

The compilation of the sequence data thus obtained reveals that in 3', the main part of the isolated sequence corresponds to the htfIIIA sequence of DREW et al. In 5', the longer sequence starts in position 80 of the htfIIIA sequence of Arakawa et al., shown in Figure 2F, and reveals the insertion of a C nucleotide in position 127 in relation to this sequence. If it can be supposed that the synthesis of the cDNA in the application of the technique described above is not complete, the insertion of a nucleotide nevertheless creates a major problem. In fact, the addition of a nucleotide in the coding sequence creates a shift in the reading frame. In order to verify the presence of this nucleotide in the htFIIIA gene, human genomic DNA was analysed by PCR. This DNA was subjected to a PCR reaction using pfu polymerase (STRATAGENE) or Taq polymerase (Perkin Elmer) using the oligonucleotides OLT5 and TFIIIA SEQ2 called SEQ ID N°5 and SEQ ID N°7 respectively as primer. The two PCR products were cloned (TA cloning Kit) then sequenced.

Analysis of the sequence data confirms the presence of this additional C nucleotide in relation to the Arakawa htfIIIA sequence for these two amplifications. The ATG initially described as start codon of proteinic synthesis for Arakawa htfIIIA can therefore no longer be considered as such.

The assembly of 5' and 3' sequences is then carried out and a unique plasmid containing the sought hTFIIIA sequence of the present invention is obtained. The complete hTFIIIA coding sequence is restored in the following manner. A clone originating from the amplification of the genomic DNA is digested using the restriction enzymes EcoRI and HindIII, and after purification, a fragment of approximately 350 bp is obtained. Furthermore, a clone originating from the

amplification of the 3' part using the restriction enzymes HindIII and SmaI is digested and after purification, a fragment of approximately 930 bp is obtained.

The ligation of these fragments in the plasmid pYX223 (expression vector for the yeast - R§D) previously digested by EcoRI and SmaI is then carried out.

A detailed account of the conditions under which the operations indicated above can be carried out is given below in the experimental part. A plasmid is thus obtained in which the gene of the present invention is inserted and this plasmid introduced into a host cell is also thus obtained by operating according to the usual techniques known to a person skilled in the art.

The polypeptide of the present invention can be obtained by expression in a host cell containing the DNA sequence coding for the polypeptide of the invention preceded by a suitable promoter sequence. The host cell can be a procaryotic cell, for example E. coli or a eucaryotic cell such as yeasts, such as for example ascomycetes amongst which are Saccharomyces cerevisiae or also mammalian cells such as Cos. cells

A particular subject of the present invention is an expression vector containing a DNA sequence as defined above.

Thus, such an expression vector according to the present invention contains a DNA sequence which can be the nucleotide sequence SEQ ID N°3 or the sequence beginning at nucleotide 176 and terminating at nucleotide 1270 of SEQ ID N°3. Such an expression vector according to the present invention can also contain the DNA sequences which hybridize with the sequences defined above, and/or show a significant homology with these sequences or fragments of them.

Such an expression vector according to the present invention can also contain DNA sequences which comprise modifications introduced by suppression, insertion and/or substitution of at least one nucleotide coding for a protein with the same biological activity as the human transcription factor hTFIIIA.

Expression vectors are vectors allowing the expression

of the protein under the control of an appropriate promoter. Such a vector can be a plasmid, a cosmid or viral DNA. For the procaryotic cells, the promoter can be for example the lac promoter, trp promoter, tac promoter, β-lactamase promoter or PL promoter. For yeast cells, the promoter can be for example PGK promoter or GAL promoter. For mammalian cells, the promoter can for example be SV40 promoter or adenovirus promoters. Baculovirus type vectors can also be used for expression in insect cells.

The host cells are for example procaryotic cells or eucaryotic cells. The procaryotic cells are for example E. coli, Bacillus or Streptomyces. The eucaryotic host cells comprise yeasts as well as cells of higher organisms, for example mammalian or insect cells. The mammalian cells are for example fibroblasts such as CHO or BHK hamster cells and Cos monkey cells. The insect cells are for example SF9 cells.

The present invention therefore relates to a process which comprises the expression of the htFIIIA protein in a host cell transformed by a DNA coding for the polypeptide sequence corresponding to sequence SEQ ID N°2.

For the implementation of the present invention, the vectors used can for example be pGEX or bpAD and the host cell can be E. coli or for example the vector pYX223, and the host cell can also be S. cerevisiae.

A particular subject of the present invention is a host cell transformed with a vector as defined above, containing the htfIIIA gene according to the present invention.

A very precise subject of the present invention is the plasmid deposited at the CNCM under the number I-2071. It thus concerns the XL1-Blue/bpShtfc2LHA strain containing the htfIIIA gene according to the present invention. The operating conditions in which the present invention was carried out are described below in the experimental part. The hTFIIIA protein coded by the htfIIIA gene is therefore a transcription regulation factor. In fact, the hTFIIIA protein coded by the gene of the present invention has a biological role as a protein binding to the DNA and the

product of this gene is useful as transcription regulation factor.

In particular, the gene of the present invention is expressed in different tissues and probably plays an important role in the initiation of the transcription of the 5S ribosomal RNA gene, and in maintaining the stability of the transcription of other genes in particular involved in control functions. A very large number of diseases accompanying a transcription control disorder have recently been brought to light. It has therefore been noted that certain oncogenic products act as transcription regulation factors and can lead to canceration of cells such as for example in certain leukaemias or also that production of the regulation factor Hox2-4 in too great a quantity induces leukaemia in mice.

Furthermore, in some hereditary diseases, the protein concerned can in itself be normal, the pathogenicity results from the transcription mechanism of the gene coding for this protein. In particular, many hereditary diseases show an abnormality in the quantity of proteins synthesized which is probably due to a disorder in proteinic synthesis which can in particular bring into play the htfIIIA gene and the coded protein as factors involved in the control of the transcription of 5S RNA.

The gene of the present invention can thus be used for the research into abnormalities in the transcription of genes, and in particular in the identification of hereditary diseases for the study of diseases implicating regulation factors and in particular the protein coded by htfIIIA.

The gene of the present invention can also be used for the treatment of certain diseases through transcription control or in the analysis of the pathogenics of these diseases.

The present invention therefore envisages the use of the htfIIIA gene of the present invention and the hTFIIIA protein of the present invention to contribute in particular to the understanding of the transcription mechanism in human beings and also to contribute to the understanding, in the diagnosis and treatment of diseases linked to a disturbance in the

transcription mechanism. Thus hTfIIIA and the htFIIIA protein could be used in the diagnosis or identification of hereditary diseases such as certain cancers or of other diseases resulting from abnormal transcription control. These factors can also be useful in the analysis of the transcription regulation mechanisms.

Therefore a subject of the present invention is the use of the DNA sequence of the gene of the human transcription factor htfIIIA or of the polypeptide having the function of human transcription factor coded by the said DNA sequence as it is defined above, for the preparation of compositions useful in the diagnosis or treatment of diseases linked to a disorder in transcription control.

Such compositions are prepared under the usual conditions known to a person skilled in the art.

A more precise subject of the present invention is the use as defined above in which the disease concerned is cancer. Figures 1 to 5 below show the following illustrations. Figure 1 represents the comparison of the hTFIIIA protein of the present invention with the DREW hTFIIIA protein.

Figure 2 represents the comparison of the hTFIIIA protein of the present invention with the ARAKAWA hTFIIIA protein.

Figure 3 represents the comparison of the DREW hTFIIIA protein with the ARAKAWA hTFIIIA protein.

Figure 4 represents the DREW htfIIIA sequence and the corresponding hTFIIIA protein.

Figure 5 represents the ARAKAWA htfIIIA sequence and the corresponding hTFIIIA protein.

The sequences indicated in the present invention i.e.: SEQ ID N°1 to SEQ ID N°10 are described below.

The experimental part below allows the description of the present invention without however limiting it.

Experimental part

Example 1: cloning and sequencing of the hTFIIIA gene

I) Extraction of total RNA originating from the RAJI human cell line (RNA Plus, BIOPROBE)

The RAJI human cell line was chosen as a source of total RNA.

The RAJI cells used were cultured under the usual culture conditions for this line known to a person skilled in the art.

To extract the total RNA of these cells a standard protocol is carried out using RNA Plus ® (BIOPROBE SYSTEMS) commercial extraction solution.

Then the following is carried out:

a) homogenization:

The cells cultured in suspension are pelleted without being washed beforehand in order to avoid the risk of degradation of the mRNA then are lysed by adding the extraction solution of the RNA Plus ® kit at a rate of 6 ml per 10^7 cells. The samples of homogenate obtained can be stored at - 70 °C.

b) extraction of the RNA:

After homogenization, the homogenate obtained in a) above is left at 4°C for 5 minutes in order to allow the complete disassociation of the nucleoproteinic complexes then 0.2ml of chloroform per 1ml of the RNA Plus ® solution is added, as above in a), the medium is agitated vigorously for 15 seconds and left to rest in ice for 5 minutes, followed by centrifuging at 12000 g and at 4°C, for 15 minutes.

Two clearly visible phases then form: the DNA and the proteins are found in the organic phase (lower phase) and at the interface. The RNA is in the aqueous phase (upper phase) which represents approximately 40 to 50 % of the total volume.

c) Precipitation of the RNA:

The aqueous phase obtained in b) is transferred into a new tube, a volume of isopropanol is added and the sample is placed at 4°C for 15 minutes, followed by centrifuging for 15 minutes at 4°C and at 1200 g. A precipitate is obtained which forms a yellow-white pellet at the bottom of the tube.

d) Washing the RNA:

The supernatant of the solution obtained in c) is eliminated then the pellet is washed with a 75 % ethanol solution using at least 0.8 ml of ethanol per 50 to 100 micrograms of RNA. The medium is mixed (vortex), centrifuged for 10 minutes at 7500 g at 4°C and dried under vacuum. The RNA obtained is

then taken up in 60 microlitres of Tris 10 mM EDTA 1 mM pH=7.5.

II) Synthesis of cDNA

a) Reagents used:

The commercial kit Gene Amp® RNA PCR Kit (Perken Elmer) was used for this cDNA synthesis.

By using this kit, the reverse transcription of RNA to cDNA is firstly obtained by reverse transcriptase MuLV (Murine Leukaemia Virus). An RNase inhibitor isolated from human placenta is included in order to inhibit certain mammalian RNases. The fragments of cDNA are amplified by polymerase chain reaction (PCR). The enzyme used for this reaction is pfu polymerase (Stratagene).

The term dNTP designates the dGTP, dATP, dTTP and dCTP nucleotides.

The term PCR Buffer designates the solution containing 500 mM KCl and 100 mM HCl at pH 8.3.

The term Oligod(T)16 designates a nucleotide sequence constituted by 16 dTTP nucleotides.

Oligonucleotides are used as primers in the technique described below.

The concentrations indicated below represent the final concentrations in the reaction medium.

b) Synthesis of the cDNA by reverse transcription:

2 microlitres of the total RNA (1 microgram) obtained above in 1)d) are pre-incubated at 65°C for 5 minutes, then 8 microlitres of the following reaction solution: 5mM MgCl2, 1xPCR buffer, 1 mM of each dNTP, 5 % of DMSO, 1 U/microlitres of RNase inhibitor, 2.5 U/microlitres of reverse transcriptase MuLV, 2.5 microlitres of oligo(dT)16 is added. The solution is then incubated at 42°C for one hour, then at 99°C for 5 minutes then at 5°C for 5 minutes.

III) Amplification by PCR, cloning and sequencing of the 3' and 5' nucleotide sequences

a) Reaction conditions:

Escherichia coli (E. coli) XL1- Blue type K12 (Stratagene) bacteria was used for the preparation of the plasmids of the present invention.

Growth of this bacteria was carried out according to the usual conditions in LB liquid medium which contains 10 g of bactotryptone, 5 g of yeast extract and 10 g of NaCl per litre of water and which also contains 100 microg/ml of ampicillin (SIGMA).

The colony was removed onto a solid LB + agar + ampicillin medium then cultured in 100 ml of LB medium and incubated to OD (600nm) = 0.8.

The incubation was carried out at 37°C under a normal atmosphere and agitation at 225 rpm.

The viability of the strain is verified when the strain grows on LB + ampicillin medium at 100 microg/ml, the insert containing a gene for resistance to ampicillin bla.

It can be noted that a gene for resistance to ampicillin bla is part of the vector of the kit (TA cloning Kit - Invitrogen) in which the fragments of htfIIIA are cloned. Thus, selection of strains containing the plasmids containing the htfIIIA gene of the present invention can be carried out by culture of the strains in this medium which contains ampicillin (100 microg/ml), such a medium allowing the survival only of strains which contain the gene for resistance to ampicillin and therefore only strains which contain the htfIIIA gene of the present invention.

For the preservation of the strains obtained, 15 % glycerol is added to the culture medium: the cultures are therefore preserved in the LB + 100 micrograms/ml of ampicillin + 15 % of glycerol at the bacterial concentration of OD (600nm) = 0.8 suspension medium in the form of aliquots in cryotubes of 1 ml per tube.

For the sequencing, the plasmid DNA of several bacteria originating from each of the cloning procedures indicated below is prepared using a commercial kit (Qiagen Plasmids kit). The fragments corresponding to the htfIII coding sequence are sequenced on the two strands according to standard techniques known to a person skilled in the art (use of the sequencer ABI 377 XL, Perkin Elmer)

b) <u>Amplification by PCR, and cloning of the 3' and 5' nucleotide sequences</u>:

1) Amplification and cloning of the 3' nucleotide sequence
Two amplification primers (primers) were chosen according to the published ARAKAWA HTfIIIA sequence. These OLT3 or TFIIIA3'SmaI and OLT5.2 primers are called SEQ ID N°10 and SEQ ID N°6 respectively.
These oligonucleotides are chosen from the hTFIIIA sequence published by ARAKAWA (Figure 5) and are synthesized according to standard methods known to a person skilled in the art.
The TFIIIA3'SmaI oligonucleotide introduces a restriction site SmaI downstream of the coding sequence. This site will allow the fusion of the htfIIIA 3' nucleotide sequence with a coding sequence for the hemaglutinine TAG peptide.
Thus, the peptide resulting from the expression of the cloned sequence will therefore consist of both the htfIIIA sequence of the present invention and that of TAG HA and can therefore be detected by Western analysis according to usual techniques known to a person skilled in the art.
The following process is then carried out: 2 microlitres of cDNA obtained above in II) b) is added to 50 microlitres of the following reaction solution: 2mM MgCl2, 1xPCR buffer, 200 nanograms/ml of each dNTP, the TFIIIA3'SMAI and OLT5.2 primers at a rate of 0.15 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNA polymerase.
The cDNA is thus subjected to 30 PCR cycles firstly at 94°C , for one minute then at 65°C for 1 minute then at 72°C for 1 minute.
The products amplified by PCR thus obtained are therefore 3' fragments of approximately 970 base pairs.
The 3' fragments obtained above are cloned in the pCRII vector using the TA cloning Kit (Invitrogen)
The plasmid thus obtained is called 5.2 Raji 2.9.
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid 5.2 Raji 2.9 is thus obtained.
2) Amplification and cloning of the 5' nucleotide sequence
The 5' portion of the htfIIIA gene of the present invention was isolated using the said 5' anchored PCR technique using a

commercial kit (5'RACE System, Rapid Amplification of cDNA Ends, GIBCO BRL).

Two amplification primers (primers) were chosen from the published ARAKAWA htfIIIA sequence (cf. Figure 5).

These TFIIIAPCR5' and TFIIIA SEQ2 primers are called SEQ ID N°8 and SEQ ID N°7 respectively.

A homopolymeric chain is added to the 3' end of the cDNA using dATP and terminal deoxynucleotidyl transferase (TdT): 10 microlitres of cDNA obtained above in II) b) are incubated at 37°C for 10 minutes in the 1 X tailing buffer reaction solution (Commercial kit solution) and 0.2 mM of dATP and TdT. The TdT is deactivated for 10 minutes at 65°C and the reaction is then brought to 4°C.

The reaction is then directly amplified by PCR: 10 microlitres of the TdT reaction are added to 50 microlitres of PCR reaction solution i.e. 1.5 mM of MgCl2, 1xPCR buffer, 200 nanomoles/ml of each dNTP, UAP and TFIIIA PCR5' primers at a rate of 0.2 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNApolymerase.

The UAP primer is an oligonucleotide provided with the commercial kit.

The cDNA is thus subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 65°C for 1 minute then at 72°C for 1 minute.

The products amplified by this first PCR i.e. PCR1 are subjected to a second amplification reaction by PCR using the UAP primer and a specific TFIIIASEQ 2 primer. The following process is carried out: 5 microlitres of PCR1 are added to 50 microlitres of the PCR reaction solution indicated below (1.5 mM of MgCl2, 1xPCR buffer, 200 micromoles/l of each dNTP, the UAP and TFIIIA SEQ2 primers at a rate of 0.2 micromoles/l for each, 5 % DMSO and 2.5 U AmpliTaq DNA polymerase.

The DNA is then subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 65°C for 1 minute then at 72°C for 1 minute.

The products amplified by this second PCR i.e. PCR2 are purified on agarose gel. The 5' fragments of approximately 380 base pairs are thus isolated.

The 5' fragments obtained above are thus cloned in the pCRII vector using the TA cloning Kit (Invitrogen).
The plasmid thus obtained is called cDNA-DMSO-3
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid cDNA-DMSO-3 is thus obtained.

3) Verification of the 5' sequence by amplification of the genomic DNA - Construction of the 5 geno-3 plasmid
Human genomic DNA is extracted from human liver cells according to the usual methods known to a person skilled in the art.
Amplification by PCR of the human genomic DNA is carried out in the following manner:
2 micrograms of human genomic DNA obtained as indicated above is added to 100 microlitres of the following PCR reaction solution: 2mM MgCl2, 1 x native Pfu DNA polymerase buffer, 200 nanograms/ml of each dNTP, the OLT5 and TFIIIA SEQ2 primers at a rate of 0.15 micromoles/l for each, 5 % DMSO and 5 U pfu polymerase.
OLT5 and TFIIIA SEQ2 are called SEQ ID N°5 and SEQ ID N°7 respectively.
The reaction medium is thus subjected to 30 PCR cycles, firstly at 94°C for one minute, then at 60°C for 1 minute, then at 72°C for 1 minute.
The products amplified by PCR thus obtained are fragments of DNA of approximately 360 base pairs.
The fragments thus obtained are cloned in the pCR-Script vector using the pCR-Script SK(+) Cloning kit (Stratagene).
The plasmid thus obtained is called 5 geno-3.
This plasmid is transferred into the XL1 Blue E. coli strain.
The E. coli strain transformed by the plasmid 5 geno-3 is thus obtained.

4) Cloning of the htfIIIA gene according to the present invention.
Construction of the pYX TFIIIALHA plasmid
The complete htfIIIA coding sequence is restored by assembly of the two 3' and 5' fragments obtained above in III) b)1) and III) b)3).

A Hind III restriction site located on each of the 3' and 5' fragments obtained above makes it possible to restore the complete sequence.

The 5 geno-3 plasmid obtained above in III) b)3) is digested by the EcoR1 and HindIII restriction enzymes.

The EcoR1 site is located 11 nucleotides upstream of the coding sequence.

Fragments of approximately 350 base pairs are obtained after purification on agarose gel.

Ligation with the vector pYX/EcoRI + HindIII is then carried out and the vector pYXTFIIIA5' is obtained.

The addition of the 3' fragment to the 5' fragment is then carried out: the 5.2 Raji 2.9 plasmid obtained above in III) b)1), is digested by the restriction enzymes HindIII and SmaI.

After purification on agarose gel, a fragment of approximately 930 base pairs is obtained. This fragment is inserted into the pYXTFIIIA5' plasmid obtained above, previously digested by the restriction enzymes SmaI and HindIII.

The pYXTFIIIALHA plasmid is thus obtained which therefore contains the hTFIIIA gene of the present invention.

Example 2: Construction of the XL1 Blue/pYX TFIIIALHA strain

The preparation of the XL1-Blue/ pYX TFIIIALHA strain, is carried out according to techniques known to a person skilled in the art (ref above: Sambrook, Fritsh and Maniatis) from the XL1- Blue type K12 E. coli strain (Stratagene), and the pYX TFIIIALHA plasmid obtained above in Example 1 is introduced.

Example 3: Construction of the bpS-tfC2LHA plasmid

The vector bpS-SK+ (Stratagene) is used, in which an insert coding for the htFIIIA gene of the present invention is integrated. The following process is carried out: the pYXTFIIIALHA plasmid obtained above in Example 1 is digested by the restriction enzyme EcoRI, this end is filled using DNA Polymerase (Klenow fragment) in the presence of dNTP. This plasmid is then digested by Nhe I and the fragment corresponding to the htfIIIA sequence according to the

present invention is purified. This fragment is inserted into the bpS-SK+ vector prepared as follows: the vector is digested by EcoRI, this site is filled using DNA polymerase then digested by XbaI.

The plasmid bpS-tfC2LHA is thus obtained.

Example 4: Construction of the XL1-Blue/bpS-tfC2LHA strain

For the preparation of the XL1-Blue/bpS-tfC2LHA strain, techniques known to a person skilled in the art, using XL1-Blue type K12 E. coli strain (Stratagene) are carried out, and the bpS-tfC2LHA plasmid obtained above in Example 3 is introduced.

A sample of the strain obtained i.e. XL1- Blue type K12 E. Coli (Stratagene) containing the bpS-SK+ vector (Stratagene) with an insert coding for tfC2 (cDNA coding part containing the htfIIIA coding region) i.e. XL1-Blue/bps-tfC2LHA coding region was deposited at L'Institut Pasteur 25, rue du Docteur ROUX Paris 75015 at the CNCM on the 15th September 1998 under the number I-2071.

Example 5: Identification of the start codon of proteinic synthesis.

Purification of the hTFIII protein was described by Moorefield et al (1994) [reference: the Journal of Biological Chemistry, Vol. 269, N° 33, pp. 20857-20865, 1994, Purification and Characterization of Human Transcription Factor IIIA, B. Moorefield and R. G. Roeder].

The hTFIIIA protein identified by Moorefield has a molecular weight of 42 kDa. It can be noted that the theoretical molecular weight of the htFIIIA protein coded by the Arakawa htfIIIA sequence is 47 kDa.

Proteinic synthesis is generally started at an ATG codon. However the htfIIIA coding sequence of the present invention does not contain the ATG codon in phase.

It has been demonstrated that the different ATG codons, in particular the CTG or GTG codons are start codons of translation in natural cellular transcripts.

With techniques known to a person skilled in the art such as translation experiments in vitro with the htfIIIA sequence according to the invention obtained above in Example 1, and

by expression tests in mammalian cells such as Cos cells, the start codon of hTFIIIA proteinic synthesis according to the present invention was demonstrated.

Within the scope of the present invention, it has thus been demonstrated that the start codon of htfIIIA proteinic synthesis according to the present invention is the CTG codon which is found in position 176-178 of SEQ ID N°3.

Analysis of the results

Analysis of the results obtained by the preparations of the examples indicated above reveal the following points relating to the htfIIIA coding sequence:

- in 3' (above in III) b)1)) the main part of the sequence isolated in the present Application corresponds to the DREW htfIIIA sequence
- in 5' (above in III) b)3)) the longest sequence of fragments obtained by the preparation described above in III) b)3) begins in position 20 of the ARAKAWA htfIIIA sequence and reveals the insertion of a nucleotide in position 127 of the ARAKAWA htfIIIA sequence.

The results obtained by the preparations of htfIIIA described above according to the present invention confirm that omission of a nucleotide in position 127 in the ARAKAWA sequence really does exist in the human htfIIIA gene.